Design and development of machine translation system for Maithili to English language

A Synopsis Submitted to Mandsaur University for Registration of Doctoral Programme

Jan 2020

Submitted by

Prabhat Kumar Singh

(19ENG6CSE1103)

Supervisor

Dr. Pradeep Laxkar

Associate Professor

Dept. Of CSE

Mandsaur University , Mandsaur

MANDSAUR UNIVERSITY

Approved by UGC, Rewas Dewda Road, SH - 31, Mandsaur, Madhya Pradesh- 458001

Ph. No. +919752122999, +919425924111

Website: www.mandsauruniversity.edu.in/

Email : info@mandsauruniversity.edu.in, phd@mandsauruniversity.edu.in

MANDSAUR UNIVERSITY, MANDSAUR

Rewas Dewda Road, SH - 31, Mandsaur, Madhya Pradesh- 458001
Ph. No. +919752122999, +919425924111
Website: www.mandsauruniversity.edu.in/
Email : info@mandsauruniversity.edu.in, phd@mandsauruniversity.edu.in

Title : **Design and development of machine translation system for Maithili to English language**

Name of Candidate : **Prabhat Kumar Singh**

Project Work : **Doctor of Philosophy (CSE)**

Supervisor : **Dr. Pradeep Laxkar**

Date : **18/08/2020**

Signature of Candidate

Name of Supervisor with Complete Designation: Dr. Pradeep Laxkar

Associate Professor
Dept. Of CSE
Mandsaur University, Mandsaur

Forwarded by Supervisor:

MANDSAUR UNIVERSITY, MANDSAUR

Rewas Dewda Road, SH - 31, Mandsaur, Madhya Pradesh- 458001
Ph. No. +919752122999, +919425924111
Website: www.mandsauruniversity.edu.in/
Email : info@mandsauruniversity.edu.in, phd@mandsauruniversity.edu.in

CERTIFICATE

This is certified that the proposed Research work of the Synopsis of Ph. D thesis Entitled "Design and development of machine translation system for Maithili to English language" has not been done earlier by anybody in any University.

(Research Scholar)
Name : Prabhat Kumar Singh
Subject:CSE
Date : 18/08/2020

(Supervisor)
Name : Dr. Pradeep Laxkar
Associate Professor
Dept. of CSE
Mandsaur University, Mandsaur

ABSTRACT

One of the significant regions of Natural language processing(NLP) is Machine translation. It is used to break the language hurdle among the people through transforming one language into required language.Lots of research has been done toward designing and developing automatic translation systems for Indian languages like Gujarati ,Hindi, Tamil, Bengali , Punjabi, Sindhi, Urdu, Kannada, Marathi and have gained satisfactory results but a little research has been started for Maithili. Maithili belongs to Indo-European language family and comes under the Indo-Aryan branch.It is used by 34 million people of Bihar, Jharkhand, Odisha , West Bengal and Nepal. Going through a measurable literature review on transforming machine translation systems for Maithili to English, it is observed that there are reasonable research gaps and we decided some objectives for this purpose to fulfil the gap.We are going to propose a machine system for transforming Maithili language to English language.

Keywords: Machine translation, Natural language processing, Maithili ,Corpus

OUTLINE OF THE PROPOSED RESEARCH WORK

I. Title of the Research Work: Design and development of machine translation system for Maithili to English language

II. Importance/ Rationale of Proposed Investigation: Machine translation is one of the important areas of Natural language processing (NLP). It breaks language barrier among the people from different region by translating source language into target language. Lots of work have been done in the direction of designing and developing machine translation system for Indian languages like Hindi, Tamil, Bengali , Punjabi, Sindhi, Urdu, Kannada, Gujarati , Marathi and have gained satisfactory results but very few work has been started for Maithili. Maithili is an Indo-Aryan branch of the Indo-European language family spoken by 34 million people of northern India and Nepal. The term Maithili originates from Mithila an autonomous Indian state in old occasions . In 2003, Maithili was remembered for the Eighth Schedule of the Indian Constitution as a perceived Indian language. In March 2018 it gained the second official language status in the Indian state of Jharkhand. It is second most popular language of Nepal spoken by 11.7 % of the population. Earlier Mithilakshar was used as the script for written Maithili, but Devanagari has been primarily used since the 20^{th} century. In 2007, it was incorporated in the Interim constitution of Nepal 2063, part 1 and section 5 as a Nepalese language. There is eager need of developing NLP tools and machine translation system for Maithili.

III. Scope of Proposed Investigation: Machine translation is not just the phenomenon in which we have to replace source language words by target language words , but the application of complex linguistic knowledge like morphology, grammar, meaning, all thing is taken into consideration. There are mainly four approaches of machine translation.

1. Rule based machine translation: It is further classified into direct translation, transfer based translation and interlingual translation. In this approach linguistic knowledge gathered from linguists and on the basis of that knowledge rules are written.
2. Statistical machine translation: It is better than RBMT in the sense that there is no requirement of manual development of linguistic rules which is very costly. It is based on the analysis of bilingual corpus and uses a statistical model to generate translations. Statistical machine translation is further divided into Word Based SMT , Phrase Based SMT, Syntax Based SMT and Hierarchical SMT.
3. Example based machine translation: In Example based machine translation the translation of new input text is gathered from the reuse of already existing translations.

4. Hybrid machine translation: In this approach more than one MT methods are used in a single system.

   After designing machine translation system most important thing is to evaluate translations done by the system. Machine translation community uses Human Evaluation, BLEU, NIST or METEOR to evaluate the machine translation.

Maithili is a low resource language in terms of language technology development and it is not rich on the web also. Maithili is morphologically rich language. The first grammar of Maithili was written by George Abraham Grienson in 1881.It is a highly inflectional language and having the scope of generating many words from a single noun, adjective or verb.

Example:

Noun inflection in Maithili for Gender.

Masculine		Feminine
बाघ	tiger	बाघिन
दास	servent	दासिन
मास्टर	teacher	मास्टरनी

Adjective inflection

Masculine	Feminine
निकहा	निकही
जरलाहा	जरलाही

Verb inflection

चलू Person inflection चलह

जाई Person inflection जो

Government of India TDIL developed machine translation system for some Indian languages but there is no such system for Maithili. Few researcher starts working on development of NLP tools but there is huge scope of developing different tools like Stemmer/Lemmatizer , Parser, Word sense disambiguation, morphological analyzer, e.t.c.

Natural Language Processing: NLP is a branch of Artificial Intelligence that deals with various kinds of methods and applications of language theory. It can be classified as

1. Natural Language Understanding : Computer understanding of human language is called Natural Language Understanding
2. Natural Language Generation: It is the developing methods that will allow a computer to respond.

Basic Text Processing Tasks in NLP : Tokenization ,Stemming/Lemmatization , Spelling corrections are the basic text processing tasks.

Tokenization: Tokenization is the process of segmenting a string of characters into words. NLTK Toolkit, Stanford Core NLP , Unix code are used for tokenization.

Sentence Segmentation: It is the problem of deciding where the sentence begins and end.

Lemmatization: It is the process of reducing inflections or variants to its base form.

Am , are , is reduced to be

Car , cars , car's , cars' reduced to car

Stemming: It is the process of reducing terms to their stems, used in information retrieval.

automate, automation, automatic reduced to automat

Morphology: Morphology studies the internal structure of words, how words are built up from smaller meaningful units called morphemes.

cats built from 'cat' and 's'

There are two types of morphology inflectional morphology and derivational morphology.

Inflectional morphology: Creates new forms of the same word.

Ex. bring , brought , bringing , brings , all are from same word bring.

Derivational morphology: Creates new words by changing part-of-speech .

Ex. logical , illogical , logician , illogicality , all derived from the same word ' logic'

Parts of speech tagging: It is the process of identifying parts of speech for the given word. Noun , Pronoun ,Adjective, Verb , Adverb , Preposition , Conjunction , Determiner are the POS.

IV Location:

a) Organization/Department where the work is to be done : Computer Science and Engineering , Mandsaur University , Mandsaur
b) Geographical Area of Investigation, if any : Bihar , Nepal

V Research gaps identified in the proposed field of investigation:

1. Maithili is low resource language, before going to develop NLP tools for the language we have to create resources for the language.
2. Need of developing NLP systems like parser , morphological analyzer , translation system for Maithili language.
3. Maithili is morphologically rich language, need to find solution to handle morphologically richness and ambiguity.
4. Although some work has been done on morphology but an efficient morphological analyzer needed which can handle inflectional as well as derivational category for Maithili.
5. The highest accuracy achieved for first Maithili POS tagger was 85.88 % ,need of improving accuracy using sufficient training data.
6. The rule based first stemmer in Maithili provides 78.5 % accuracy, which can be improved.

VI Objective of the proposed study:

1. To design and develop an efficient inflectional and derivational morphological analyzer for Maithili language.
2. To design and develop first Maithili language parser.
3. To develop an algorithm to resolve word sense disambiguation in Maithili language.
4. To develop machine translation system for Maithili to English language.

VII Review of Work done on the Subject:

1. **Hajic et al.(1987)** developed a system in which they adopted transfer- like translation scheme (they did not use any intermediate pilot language), is referred as direct method. The translation process itself was carried out in batch, i.e there was no human intervention. The results obtained high quality which only required usual post editing; no human pre edition was entertained in the system.[6]
2. **Hajič et al.(2000)** used example of RUSLAN which is transfer based MT for Czech-Russian pair .Author proposed a MT system CESILKO for Czech and Slovak pair.[7]
3. **Hettige et al.(2007)** developed morphological analyzer of Sinhala language which can be used in machine translation for English to Sinhala language. Morphological analyzer has been developed with three associated dictionaries- Base dictionary, Rule dictionary, and Concept dictionary.[8]
4. **Shrivastava et al.(2008)** presented a basic HMM based Part of speech tagger in which a naïve , longest suffix matching stemmer is utilized as a preprocessor to accomplish a sensibly descent precision of 93.12 %. In this technique no prerequisite of any linguistic asset isolated from a rundown of possible additions for the language. This method cannot be as good as using detailed morphological analysis and linguistic resources.POS tagger can accomplish generally excellent outcomes whenever given great morphological data .Cost of making these resources renders such techniques unrealistic.[26]
5. **Mohanty et al.(2008)** developed a multilingual dictionary framework which was based on wordnet SYNSETS, i.e concepts .The dictionary was considered as an M x N Matrix where M is the no. of synsets (concepts) and N is the no. of languages. Multilingual dictionary developed for 11 languages.[17]
6. **Goyal et al.(2008)** developed morphological analysis and generator tool for Hindi language using paradigm approach for windows platform having GUI. In the developed model drawbacks of existing system eliminated. Users need not to have much knowledge about computers , platforms, and any programming language. Users just need some basic computer operation knowledge for software installation and operate. It has been developed for Unicode format. Tool stores all the commonly used word forms for all Hindi root words in its database. This approach prefers time and accuracy to memory space.[5]
7. **Josan et al.(2008)** considered closely related language Punjabi-Hindi (Indo –Aryan languages) due to their similarity in context of syntax and semantics. For this similarities author choose direct translation methodologies for these language pair. Author found the better accuracy of the translation system i.e. 90.67% as compared to accuracy of previous MT system of closely related languages .[9]
8. **Mahar et al.(2010)** designed and developed tag set, lexicon and word disambiguation rules for Sindhi POS.They also designed and implemented tagging and tokenization algorithm of SPOS . They achieved accuracy of 96.28% from SPOS by using rule based POS tagger.[12]
9. **Suba et al.(2011)** presented two stemmers for Gujarati, first one based on hybrid approach performing inflectional stemming where as second one based on rule based approach performing derivational stemming. Maximum Accuracy of the inflectional stemmer was 90.7% .[28]

10. **Sugandhi et al.(2011)** developed the first system for spoken language understanding based on trained statistical model and this derived from annotated corpora .The annotation is based on simple representation of semantic of the domain, so it is relatively straightforward and they called system as hidden understanding model.[29]
11. **Mall et al.(2013)** proposed a MT system for Hindi to English language based on ruled based theory .Authors also planned do develop lexical resources like vocabulary, POS tagger.[14]
12. **Patel et al.(2013)** created rule based token mapping system GH-MAP for translation between language pair Hindi and Gujarati which are sibling languages. Translation is not performed here in the sense of linguistics, but word for word translation is performed. It requires limited linguistic effort and tools for achieving the goal. The aim of the GH-MAP was not to deliver excellent translation in the feeling of phonetics rather was created to deliver right working translation adequate to cross the language obstruction.[19]
13. **Ameta et al.(2013)** proposed the use of stemming and POS tagging for transliteration. Transliteration is a process which maps source content to the target content. They used a hybrid approach - a blend of rule – based and corpus based methodology ,where they utilized POS labeled corpus and stemming rules to assist the process of transliteration and achieved overall efficiency 93.09 %.[1]
14. **Bhatti et al. (2014)** resolved the tokenization problem for Sindhi language by a proposed model and suggested two steps, first by identifying sentence boundaries and then segregated sentences decomposed into words.[3]
15. **Kuboň et al.(2014)** compared the result of MT between closely related Czech and Slovak languages , by using RBMT and SMT approaches and found for closely related languages that not only RBMT is suitable method for MT but SMT also produces the similar transformation.[10]
16. **Mall et al.(2014)** resolved issues in parsing technique by modifying CKY algorithm in machine translation from Hindi to English sentences. For these authors designed transfer module to generate English sentences using CFG. Authors enhanced grammatical rules for resolving issues like issues of gender, case and free order in databases and algorithm improved the result by 91.4%.[13]
17. **Saini et al.(2015)** reviewed various MT approaches and existing MT systems in India .Authors found more research required in MT to overcome the language barrier faced by India. Authors found after reviewing that MT developed till date have many shortcomings in terms of rules set dictionary, methodology of translation.[25]
18. **Mall et al.(2015)** developed and evaluated POS tagging algorithm for parsing the Hindi text in Unicode format .It also verifies the Hindi text according to the correct grammar .for this authors implemented four different algorithms (i) Hindi tokenizing algorithm (ii)conversion algorithm,(iii)POS tagging and (iv) translation algorithm.[15]
19. **Kumar et al.(2015)** developed English to Hindi MT system and used example based approaches. They have trained module by taking several sample texts in English and their corresponding Hindi sentences .during translation used step wise procedure.[11]
20. **Singh et al.(2017)** presented the method to train the recurrent neural network for reordering for source to target language by using semi supervised learning method .word2vec and auto encoder used for source and target language respectively. RNN has

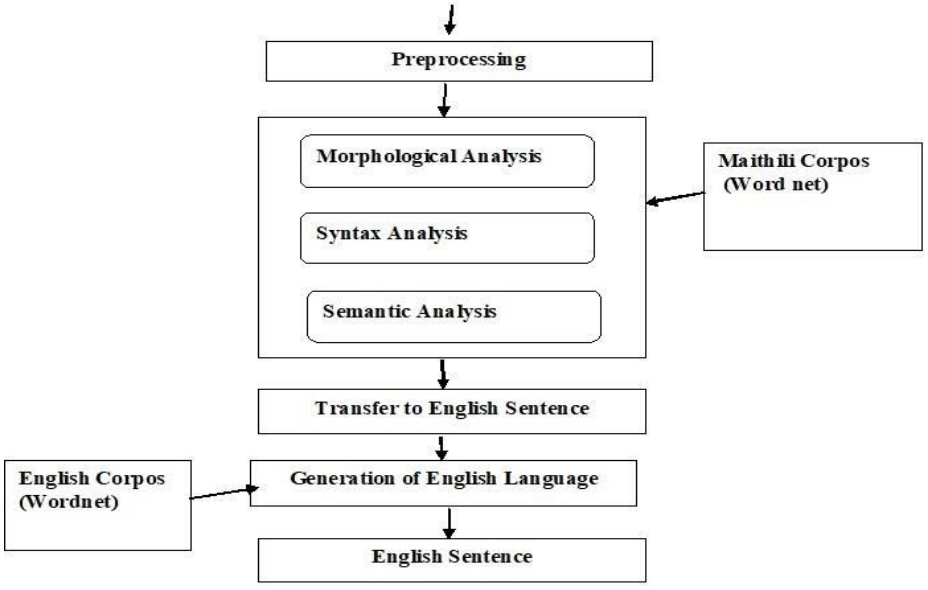
the capability to implement reordering rules on sentences. GPU solves the problem of complex computation and leads the system tow good performance because it supports massive parallel computation.[27]

21. **Mall et al.(2017)** developed an algorithm to resolve the issues in word sense disambiguation and performed correct translation from Hindi to English language. Shallow parsing model is based on Hidden Markov model where as for word sense disambiguation enhanced the modified Lesk algorithm , also compared their result with resolved issue of word sense disambiguation with the Google translator. They showed that their system resolved polysemy word but Google translator fails to resolve polysemy word.[16]
22. **Chakrawarti et al.(2017)** reviewed various Hindi – English machine translation projects and proposed possible approaches for machine translation. Authors also analyzed effectiveness of the existing online Hindi –English machine translation applications and observed that most of the application failed to produce desired output.[4]
23. **Babhulgaonkar et al.(2017)** described resent research for word alignment and language modeling problems in the translation process. They presented word based and phrase based statistical machine translation then concluded work with enlisted problems with SMT.[2]
24. **Rishikesh(2018)** build POS tagger for the Maithili language using HMM algorithm. As there were no work has been done previously for Maithili language firstly prepared a training corpus containing 10000 words and tested the system with more than 1200 known words. The accuracy measured to 95.67 % for known words.[24]
25. **Priyadarshi et al. (2019)** presented their efforts on development of POS tagger as there were no available training data to start the work firstly they developed the corpus of POS tag and defined POS tag set manually for 52190 words. Using this they have trained CRF and achieved 87.67 % accuracy with distinct feature in CRF using defined corpus. To improving the accuracy authors gathered huge raw corpus from different sources like Wikipedia dump and other Maithili web resources and, he trained neural word embedding. To generate word vector they used word 2 vec, CBOW model during the training of CRF and improved accuracy by 85.88% .[22]
26. **Priyadarshi et al. (2019)** developed first stemmer in Maithili using hybrid approach in three steps, firstly developed suffix stripping based stemmer results in some issues like under stemming and over stemming of a string which goes to many invalid strings. Then developed a POS dependent rule based approach which resulted in a decrease in the count of invalid strings. Finally to validate the system generated stems incorporated a word embedding module. Accuracy of final system was 84.6 %.[21]
27. **Pourdamghani et al.(2019)** presented a method for translation text between closely related languages without needing any parallel data. Statistical machine translation technique requires the availability of large amount of parallel data between source and target languages.[20]
28. **Nidhi et al.(2019)** discussed the resource creation efforts and divergence study for a statistical English to Maithili machine translation .Resource creation for machine translation found to be big challenge due to absence of basic tools in Maithili . Focus was on information assortment techniques, cleaning the size and structure of the content

corpora, arrangement and parallelization techniques planning, testing and examination of uniqueness between the language pair .[18]

29. **Verma et al.(2019)** showed the development of Hindi –English MT using neural approach .authors found that in the automatic evaluation BLEU metrics was producing better results for attention based model which was and improvement over the base line model.[30]
30. **Rahi et al.(2020)** presented inflectional morphological analyzer based on finite state transducer for Maithili language. According to authors this work can be seen as the essential work towards the computational development of Maithili. Conducted a series of experiments to evaluate the correctness of the system.[23]

VIII. Research methodology:



We will develop a machine translation system for the Maithili language. Steps will be as follows:

1. Input: Maithili Language
2. Preprocessing
    i. Removal of stop words
    ii. Tokenization
    iii. (iii)Stemming/Lemmatization
3. Analysis of Maithili Language
4. Transfer of Maithili Sentences into English Sentences:In the proposed work we will use statistical machine translation technique.
5. Generation of English Sentences : Finally English sentences will be generated . During this phase we also consider word sense disambiguation.

- **Hypotheses to be tested:** Our proposed system will convert Maithili sentences into English sentences with approximately 90 % accuracy.
  Example : हम हुनकर घर देखलिऐन्ह। will be converted into " I saw his house ".
- **Sources of Information:** Maithili related Website , Kaggle, Maithili literature
- **Tools and Techniques of Research:** NLP ,NLTK , TensorFlow
- **Plan of Work:** Following is the timeline for our research.

Serial No.	TASK TITLE	START DATE	DUE DATE	DURATION	% OF TASK COMPLETE	Synopsis Presentation	Data collection, Pre-processing of Data	Designing Parser , One rese	Designing word sense disa	Result analysis, Thesis wr	Result analysis, Thesis writing, Thesis submission
						Semester I	Semester II	Semester III	Semester IV	Semester V	Semester VI
						Jan-July 2020	July-Dec 2020	Jan-July 2021	July-Dec 2021	Jan-July 2022	July-Dec 2022
1	Synopsis Presentation										
1.1	Semester I	1/1/20	6/30/20	179	100%						
2	Data collection, Pre-processing of Data										
2.1	Semester II	7/1/20	12/31/20	180	0%						
3	Designing Parser , One research paper										
3.1	Semester III	1/1/21	6/30/21	179	0%						
4	Designing word sense disambiguation , One research paper										
4.1	Semester IV	7/1/21	12/31/21	180	0%						
5	Integration of tools and development of machine translation system, one research paper										
5.1	Semester V	1/1/22	6/30/22	179	0%						
6	Result analysis, Thesis writing, Thesis submission										
6.1	Semester VI	7/1/22	12/31/22	180	0%						

IX. Expected outcome of research: The proposed research lead to the development of machine translation system for Maithili-English language pair. Our system will be able to handle morphology and will be capable of handling ambiguous sentences.

X. Tentative Chapterization: Following is the tentative chapeterization

Chapter 1 : Introduction of Research

Chapter 2 : Literature Review

Chapter 3 : Methodologies and Implementation

Chapter 4 : Result and discussion

Chapter 5 : Conclusion and Future work

XI. References:

1. Ameta, J., Joshi, N., Mathur, I. (2013). Improving the Quality of Gujarati-Hindi Machine Translation through Part-of-Speech Tagging and Stemmer Assisted Transliteration. IJNLPC, 2(3), pp. 49-54.
2. Babhulgaonkar, A.R., Bharad, S. (2017). Statistical machine translation. *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM),* pp. 62-67.
3. Bhatti, Z., Ismaili, I., Soomro, W., Hakro, D. (2014) .Word Segmentation Model for Sindhi Text , American Journal of Computing Research Repository, *2*(1), pp 1-7.
4. Chakrawarti, R.K. Bansal, P. (2017). Approaches for Improving Hindi to English Machine Translation System.Indian Journal of Science and Technology, 10(16), pp. 1-8.
5. Goyal, V., Lehal, G.(2008). Hindi Morphological Analyzer and Generator. Emerging Trends in Engineering & Technology, IEEE, pp. 1156-1159.
6. Hajic, J. (1987). RUSLAN - AN MT SYSTEM BETWEEN CLOSELY RELATED LANGUAGES, Third Conference of the European Chapter of the ACL.
7. Hajič, J., Hric, J. , Kuboň V.(2000). Machine translation of very close languages, Conference: Proceedings of the sixth conference on Applied natural language processing, pp. 7-12.
8. Hettige, B. Karunananda, A.(2007). A Morphological Analyzer to Enable English to Sinhala Machine Translation, International Conference on Information and Automation , pp. 21-26.
9. Josan, G., Lehal, G. (2008). A Punjabi To Hindi Machine Translation System,Companion volume Posters and Demonstration, Manchester, pp. 157-160.
10. Kuboň, V., Vičič, J. (2014). A Comparison of MT Methods for Closely Related Languages: a Case Study on Czech - Slovak Language Pair , Association for Computational Linguistics pp . 92-98.
11. Kumar, P., Srivastava, S. , Joshi, M. (2015). Syntax directed translator for English to Hindi language. IEEE International Conference on Research in Computational Intelligence and Communication Networks, pp. 455-45.
12. Mahar, J., Memon, G.(2010). Rule Based Part of Speech Tagging of Sindhi Language, International Conference on Signal Acquisition and Processing, Bangalore , India, February 9-10, 2010,IEEE, pp.101-106.
13. Mall, S., Jaiswal, U., (2014). Resolving issues in parsing technique in machine translation from Hindi language to English language,*ICCCT*, IEEE, Allahabad,, pp. 55-58.

14. Mall, S., Jaiswal, U. (2013). Developing a system for machine translation from Hindi language to English language. 4th International Conference on Computer and Communication Technology, IEEE, pp. 79-87.
15. Mall, S., Jaiswal, U., Innovative algorithms for Parts of Speech Tagging in hindi-english machine translation language,2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Noida, 2015, pp. 709-714.
16. Mall, S., Jaiswal, U. (2017). Shallow Parsing and Word Sense Disambiguation Used for Machine Translation from Hindi to English Languages. IJIES, pp. 381-390.
17. Mohanty, R.K., Bhattacharyya, P., Kalele, S.D., Pandey, P., Sharma, A., Kopra, M. (2008). Synset Based Multilingual Dictionary: Insights, Applications and Challenges, 4th Global WordNet conference, Szeged, Hungary, January 2008, page. 321-332.
18. Nidhi, T. (2019). Machine Translation and Divergence Study for English–Maithili, Proceedings of ICT4SD 2018, pp.383-390.
19. Patel, K. ,Pareek, J. (2013). GH-MAP: translation system for sibling language pair Gujarati--Hindi. CSI Transactions on ICT, Springer.
20. Pourdamghani, N., Knight, K.(2019). Neighbors helping the poor: improving low-resource machine translation using related languages. Springer Nature,Machine Translation 33(2).
21. Priyadarshi, A., Saha, S.(2019). A Hybrid Approach to Develop the First Stemmer in Maithili,ACM transaction, FIRE'19, 2019, Kolkata, India, pp. 42-46.
22. Priyadarshi, A., Saha, S.(2019). Towards the first Maithili part of speech tagger: Resource creation and system development, Computer Speech & Language. 62.
23. Rahi, R. ,Pushp, S., Khan A., Sinha, S. (2020). A Finite State Transducer Based Morphological Analyzer of Maithili Language, arXiv:2003.00234 [cs.AI].
24. Rishikesh. (2018). Parts Of Speech Tagger for Maithili Language Using HMM, International Journal of Innovations and Advancement in Computer Science, GIET Sonipat, 7(4), pp. 206-2011.
25. Saini, S., Sahula, V. (2015). A Survey of Machine Translation Techniques and Systems for Indian Languages. Proceedings - 2015 ,ICCICT,IEEE pp. 676-681.
26. Shrivastava, M., Bhattacharyya, P. (2008). Hindi POS Tagger Using Naive Stemming : Harnessing Morphological Information Without Extensive Linguistic Knowledge,ICON-2008, 6th International Conference on NLP.
27. Singh, S., Kumar, A., Darbari, H.,Singh, L., Rastogi, A. , Jain, S. (2017). Machine translation using deep learning: An overview. International Conference on Computer, Communication and Electronics, MNIT Jaipur, pp. 162-167.
28. Suba, K., Jiandani, D., Bhattacharyya, P. (2011). Hybrid Inflectional Stemmer and Rule-based Derivational Stemmer for Gujarati. In proceedings of the 2nd Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), IJCNLP 2011, Chiang Mai, Thailand, pp.1-8.
29. Sugandhi, R., Shekhar, R., Agarwal, T., Bedi, R., Wadhai, Vijay. (2011). Issues in parsing for Machine Aided Translation from English to Hindi. Proceedings of the 2011

World Congress on Information and Communication Technologies, WICT 2011, IEEE,pp. 754-759.

30. Verma, C., Singh, A., Seal, S., Singh V., Mathur I. Charu Verma, Aarti Singh, Swagata Seal, Varsha Singh, Iti Mathur I. (2019).International Journal of Scientific and Technology Research, 8(11), pp. 2710-2714.

XII. Appendix (If any)

- **Research or Review Papers**
- **Questionnaire**

Name and Signature of Candidate : Prabhat Kumar Singh

Name & Signature of Supervisor : Dr. Pradeep Laxkar

Date: 18/08/2020

Document Information

Analyzed document	ForChecking.pdf (D82373639)
Submitted	10/22/2020 10:02:00 AM
Submitted by	Dr Pradeep Laxkar
Submitter email	drpradeep.laxkar@meu.edu.in
Similarity	4%
Analysis address	drpradeep.laxkar.meu@analysis.urkund.com

Sources included in the report

W	URL: https://link.springer.com/article/10.1007/s40012-012-0009-6 Fetched: 10/22/2020 10:03:00 AM	2
W	URL: https://www.researchgate.net/publication/299435025_A_Review_on_Machine_Translation ... Fetched: 10/22/2020 10:03:00 AM	1
W	URL: https://link.springer.com/chapter/10.1007%252F978-981-13-7166-0_37 Fetched: 10/22/2020 10:03:00 AM	1
W	URL: https://arxiv.org/abs/2003.00234 Fetched: 10/22/2020 10:03:00 AM	1
W	URL: https://www.researchgate.net/publication/249011599_Improving_the_Quality_of_Gujara ... Fetched: 6/14/2020 10:37:42 AM	1
W	URL: https://www.researchgate.net/publication/257808918_GH-MAP_translation_system_for_s ... Fetched: 10/22/2020 10:03:00 AM	1
W	URL: https://www.ijeat.org/wp-content/uploads/papers/v8i4/D6037048419.pdf Fetched: 12/13/2019 6:51:38 PM	1